

c246 Problem Set 5

Due 5 Mar 2002, 5pm by email to homework@c246.lbl.gov

1. (10 points) What are the advantages of using multiple sequence alignments over pairwise sequence alignments?
2. (10 points) What are the disadvantages of using multiple sequence alignments instead of pairwise sequence alignments?
3. (10 points) What is the major problem with progressive alignment schemes?
4. (10 points) What are the major differences between a multiple sequence alignment of genomic DNA sequences and a multiple sequence alignment of a group of homologous proteins?
5. (10 points) Explain how knowledge of the evolutionary relationships between organisms can be used to help interpret multiple sequence alignment results?
6. (10 points) Explain how one can use multiple sequence alignments to infer whether sequences are paralogs or orthologs? How does this change when there are multiple sequences from the same organisms?
7. (20 points) Extend the program from Assignment 4 to read in a FASTA file of sequences, to perform all pairwise alignments of the sequences and to perform a master-slave multiple alignment. Apply this program to sequences from GenBank with the following accession ID's:

CAA38458, AAB38383, CAC37335, AAF36979, AAB87968, AAK49948,
10948406, Q9IA06, AAD05435.1, AAF89677.1, NP_491028, NP_005622.1,
NO_036939.1

8. (20 points) Name programs that employ three different classes of algorithms for making multiple alignments. Align the above sequences using each of these programs (from web or elsewhere). Comment on the differences between the results. Score each of the above alignments using BLOSUM45, Gap penalty -8, using SP and using minimum entropy (calculate using p_{ia} 's only for observed characters).